Automated Seizure Detection using Transformer Models on Multi-Channel EEGs

Yuanda Zhu
School of Electrical and Computer Engineering
Georgia Institute of Technology
Atlanta, GA, USA
yzhu94@gatech.edu

May D. Wang
Dept. of Biomedical Engineering
Georgia Institute of Technology and Emory University
Atlanta, GA, USA
maywang@gatech.edu

***Abstract*—Epilepsy is a prevalent neurological disorder characterized by recurring seizures, affecting approximately 50 million individuals globally. Given the potential severity of the associated complications, early and accurate seizure detection is crucial. In clinical practice, scalp electroencephalograms (EEGs) are non-invasive tools widely used in seizure detection and localization, aiding in the classification of seizure types. However, manual EEG annotation is labor-intensive, costly, and suffers from low inter-rater agreement, necessitating automated approaches. To address this, we introduce a novel deep learning framework, combining a convolutional neural network (CNN) module for temporal and spatial feature extraction from multi-channel EEG data, and a transformer encoder module to capture long-term sequential information. We conduct extensive experiments on a public EEG seizure detection dataset, achieving an unweighted average F1 score of 0.731, precision of 0.724, and recall (sensitivity) of 0.744. We further replicate several EEG analysis pipelines from literature and demonstrate that our pipeline outperforms, current state-of-the-art approaches. This work provides a significant step forward in automated seizure detection. By enabling a more effective and efficient diagnostic tool, it has the potential to significantly impact clinical practice, optimizing patient care and outcomes in epilepsy treatment. Codes available on GitHub [1].**

***Index Terms*—EEGs, seizure detection, transformer model**

I. Introduction

Epilepsy, a neurological disorder distinguished by recurring seizures, affects over 3 million individuals in the United States [1] and approximately 50 million globally [2]. Among the complications associated with this condition, Sudden Unexpected Death in Epilepsy (SUDEP) presents a severe risk, claiming the lives of 1 in 1000 epilepsy patients annually [3], [4]. Given this threat, early and accurate seizure detection is pivotal in clinical practice, as timely intervention can significantly reduce mortality risk [5].

Electroencephalograms (EEGs) serve as a non-invasive tool for seizure detection in clinical practice. EEGs measure the electric potential differences on the scalp [6], helping to confirm the occurrence of a seizure, localize epileptogenic regions within the brain, and classify the type of seizure [7]. In the process of diagnosing seizures in known epileptic patients, clinicians search for characteristic discharges and patterns in pre-ictal (before a seizure), inter-ictal (between seizures), and ictal (during a seizure) periods. However, in a clinical setting, this manual annotation is labor-intensive, costly, and suffers from low inter-rater agreement [8], [9].

Over the years, significant strides have been made in the realm of automatic seizure detection, with research leveraging both traditional feature engineering and deep learning techniques. Unfortunately, the complex nature of EEG signals presents formidable challenges, leading to the underwhelming performance of existing EEG seizure detection methods. Additionally, existing approaches often find it challenging to mitigate patient-to-patient variations and learn seizure-specific features, leading to significant overfitting issues.

Hence, there is an urgent need to develop novel AI approaches that facilitate more effective, efficient, and reliable seizure detection. In this work, we propose a novel deep learning framework for automatic seizure detection. The proposed framework consists of a convolutional neural networks (CNNs) module and a transformer encoder module. The CNN module is effective in extracting both temporal and spatial features from the multi-channel EEG data. The transformer encoder module further captures the long-term sequential information from the feature vectors. Our approach achieves unweighted average F1 score of 0.731, unweighted average precision of 0.724 and unweighted average recall (sensitivity) of 0.744.

The main contributions of this work are two-fold:

- We propose a novel deep learning framework for patient-specific seizure detection. The proposed framework can effectively extract spatial and temporal seizure-specific features from multi-channel, raw EEG signals, while capturing the long-term sequential information.
- We conduct extensive experiments to evaluate the performance of the proposed pipeline on the world's largest public EEG seizure detection dataset. The results demonstrate that our pipeline outperforms competitive state-of-the-art approaches.

II. Related Works

A. Clinical Diagnosis of Seizure

The process of clinical EEG examination follows a sequence of stages: electrode placement, data acquisition, EEG

This work was supported by a Microsoft Azure Cloud grant, and the Petit Institute Faculty Fellow, and Carol Ann and David D Flanagan Faculty Fellow Fund to Prof. May D. Wang.

[1]https://github.com/UnitedHolmes/seizure_detection_EEGs_transformer_BHI_2023

recording collection, and the generation of clinical reports. The commonly adopted arrangement for scalp electrodes is the standard 10/20 system [6]. Here, electrodes are symmetrically positioned on both the left and right sides of the scalp, measuring the electrical potential across the pre-frontal, frontal, parietal, temporal, and occipital lobes.

In clinical practice, clinicians often employ video EEG tests or video EEG monitoring as reliable methods for detecting and diagnosing seizures [10]. A video EEG simultaneously captures patient behavior and movements via video, alongside brain electrical activity through the placement of scalp electrodes. This dual modality of recording provides a comprehensive perspective of the patient's physical and neurological states during seizure events.

The use of video EEG allows clinicians to ascertain whether a seizure or event corresponds to anomalous electrical activity in the brain, identify unusual EEG features, and confirm the type of seizure. Key characteristics that clinicians specifically analyze in the EEG signal include evolution, spike and wave morphology, rhythmicity, synchrony, and frequency. Each of these features offers vital insights into the nature of the seizure, enabling a more targeted approach to treatment. Through a careful examination of these parameters, clinicians can establish a more accurate diagnosis and personalized care strategy for individuals suffering from seizures.

B. *Automated EEG Seizure Detection*

EEG seizure patterns exhibit significant inter-patient variability, which can range from focal spikes in patient-specific channels to generalized spikes across all channels [11]. The task of identifying these markers is labor-intensive, subjective, and often suffers from poor inter-rater agreement with scores varying from 0.46 to 0.87 [8], [9].

In an effort to reduce subjectivity and alleviate the manual labor involved, the focus has been gradually shifting toward automated annotation systems. These systems leverage hand-engineered features such as coherence [12], entropy [13], and a range of other statistical and spectral features [14]. Prior to 2014, these handcrafted features, often referred to as 'shallow features', were typically categorized into the time domain, frequency domain, and wavelet domain [15].

Recently, there has been a burgeoning interest in utilizing deep learning systems for EEG data analysis. These systems promise the capability of automatically extracting relevant features from EEG waveforms, thereby potentially enhancing the efficiency and accuracy of seizure detection. The four principal categories of deep neural networks exploited for this purpose encompass CNNs [16]–[18], Recurrent Neural Networks (RNNs) [19]–[21], Temporal Convolutional Networks (TCNs) [22], [23], and Transformer models [24], [25].

CNN-based models, predominantly used in image processing, excel at learning local and spatial information. However, they often fall short in capturing long-term temporal dependencies [16], [17]. On the other hand, RNNs, including the Long Short-Term Memory (LSTM) models, are favored for sequence modeling tasks as they adeptly capture temporal information within sequences. Nonetheless, RNNs typically struggle to extract spatial domain information, especially the multi-channel information inherent in EEGs. Thus, RNNs and LSTM models are often utilized along with CNN models [20], [21]. Still, RNNs are prone to issues such as exploding gradients and often fail to capture long-term temporal dependencies.

More recently, TCNs [22], [23] and Transformer models [26] have emerged as promising tools for sequence modeling and text analysis. They exhibit proficiency in learning long-term temporal information, but similar to their predecessors, they may struggle to capture the spatial domain information present in multi-channel EEGs. To address this issue, Sun et al. [24] proposed a deep learning pipeline that utilized a 2D CNN model to generate features vectors from the multi-channel intracranial EEGs (iEEGs) and the transformer encoder module that learned both the temporal information within individual channels and the quantify the attention among different channels. Likewise, Li et al. [27] extracted the frequency domain features from the EEGs using short-time Fourier transform (STFT) before applying the CNNs and transformer models for seizure prediction.

As such, while each of these deep learning models has its strengths, continued research is needed to design the best framework that effectively integrates some of these models and further improves seizure detection accuracy.

III. METHODS

As shown in Figure 1, our proposed approach consists of a preprocessing module, a CNN module for feature extraction and a transformer module for classification.

A. *Data Preprocessing*

In this study, we employed a specific process for the preprocessing of EEG data. Initially, we extracted token-level EEG signals and time-stamped annotations from the TUSZ dataset. We proceeded to apply a bandpass filter in order to retain signals within the frequency range of 0.5 Hz to 100 Hz. Subsequently, two notch filters were utilized to eliminate signals at 1 Hz and 60 Hz, which are typically associated with heart rate and power line noise, respectively.

In terms of the binary classification task, seizure signals were divided into four-second segments with a 75% overlap. Conversely, all non-seizure signals were partitioned into four-second segments without any overlap.

All signals were resampled to 250 Hz if their original sampling rate deviated from 250Hz. Moreover, we identified any signal segment exceeding 500 microvolts as noise and subsequently excluded such segments from our analysis.

B. *CNN Module for Feature Extraction*

First, we use the EEGNet as the CNN module for feature extraction. EEGNet [28] is a compact CNN-based model to extract spatial and temporal features from multi-channel EEGs. EEGNet has proven effective in various motor imagery/ brain-computer interface (BCI) tasks [29], [30], as well as other

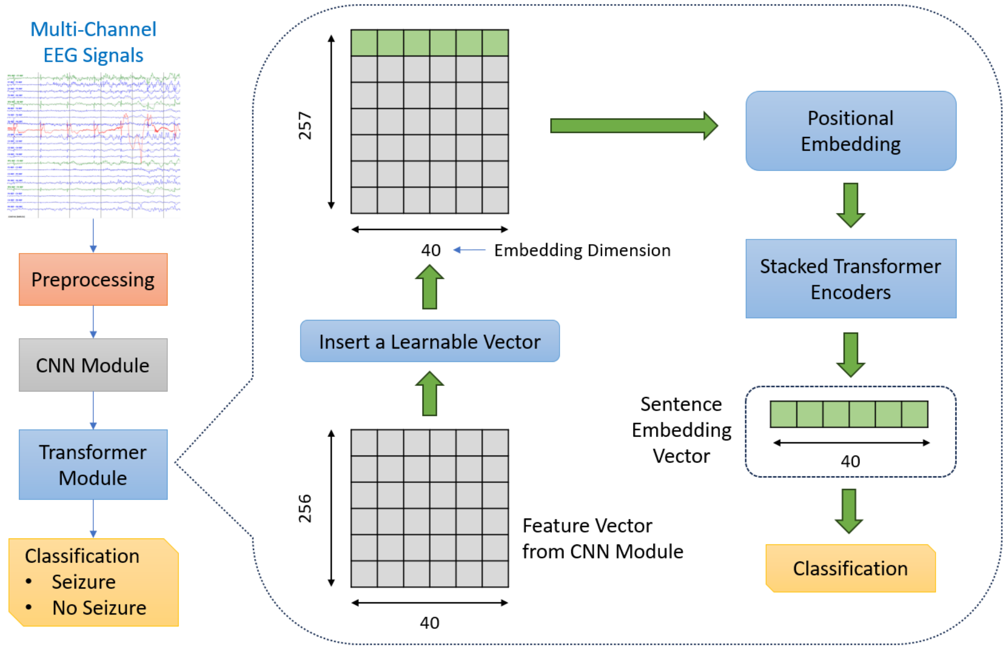

Fig. 1. Overall flowchart diagram for our proposed approach for EEG seizure detection.

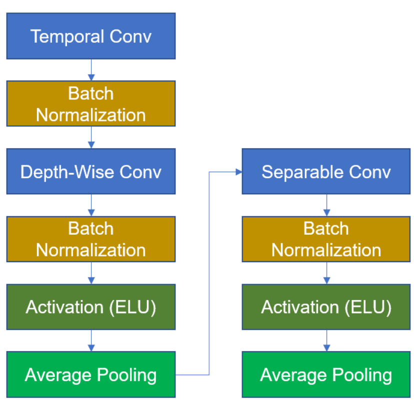

Fig. 2. Using three convolutional layers, the CNN module effectively extracts both spatial and temporal features from the raw EEGs.

EEG tasks such as seizure detection [31], [32] or seizure type classification [33].

As shown in Figure 2, the CNN module mainly consists of three convolutional layers, along with multiple batch normalization, activation, and pooling layers. The first convolutional layer is designed to extract temporal information from the multi-channel EEGs using F_1 filters with the kernel size $(1,\ K_{C1})$. Here K_{C1} is the filter size along the temporal dimension. Popular choices of K_{C1} are 64 and 32, which are approximately one-fourth or one-eighth of the sampling frequency (250 Hz), respectively. This setting allows the convolutional layer to extract temporal features above 4 Hz or 8 Hz. The resulting output of the first convolutional layer consists of F_1 temporal feature maps.

The second convolutional layer extracts spatial information from the multi-channel feature maps, utilizing $F_1 \times D$ filters with a kernel size of $(C,\ 1)$. Here, C denotes the number of EEG channels, which in this work is 22. This setup enables the convolutional layer to learn spatial features across all channels effectively. The output from this depth-wise convolutional layer comprises $F_1 \times D$ feature maps, wherein each temporal feature map generated by the previous layer corresponds to D output feature maps.

The third layer, a separable convolutional layer, independently learns a temporal summary for each feature map, and subsequently mixes these feature maps. The output feature map of separable convolution contains temporal information.

Complementing these, the batch normalization layers work to speed up and stabilize the training process, the exponential linear unit (ELU) activation layers introduce nonlinearity, and the pooling layers are employed to abstract temporal features.

C. *Transformer Model*

RNN-based models tend to have limitations when dealing with long sequences [34], which led to the proposal of the transformer model equipped with multihead self-attention (MSA) modules [26]. Transformer model can simultaneously encode and align words within a sentence, significantly enhancing semantic accuracy in learning contextualized word embeddings and improving performance in downstream tasks.

The key component of the transformer model is the self-attention module. The self-attention score is computed as a weighted sum of Value matrix (V), with coefficients derived from the dot-product of Query (Q) and Key (K) matrices.

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V$$

These matrices are generated using linear transformations, specifically, $Q = XW^Q$, $K = XW^K$, and $V = XW^V$. The weight matrices W^Q, W^K, and W^V are learnable parameters, while X denotes the latent representation from the previous module and d_k represents the embedding dimension.

In this work, we choose the Transformer model as it is effective in capturing the global relationship in a segment of signals, often more powerful than RNN and TCN models. We use 4 MSA heads and 2048 as embedding dimension.

D. Summary of the Proposed Approach

Inspired by [24], we consider each temporal feature map generated by the separable convolution layer as a word, and the length of the feature map as the embedding size. We show the key steps of the transformer module in Figure 1. Similar to the idea of learning sentence embeddings in the BERT model [35] (which inserts a special token "CLS" in the beginning of the sentence), we insert a one-dimensional, learnable vector at the beginning of the temporal feature map. This one-dimensional vector has the same length as each temporal feature map. After adding the positional embedding, we feed the entire feature map into the transformer encoder for binary classification. Detailed architecture of the proposed model is shown in Table I, along with the output data shape after each layer. Please note that tensor permutation is not included in this table.

IV. Experiments and Results

A. Dataset Description

The Temple University Hospital EEG Corpus (TUH EEG) stands as the largest publicly available dataset of EEG recordings worldwide [36]. A subset of this, known as the TUH EEG Seizure Corpus (TUSZ), is specifically designed for seizure detection, making it the largest public dataset for this purpose.

The predefined training set encompasses 1,185 EEG sessions from 592 patients, while the testing set includes 238 EEG sessions from 50 patients. The TUSZ data was collected using electrodes arranged on the scalp following the standard 10-20 format. This consistent electrode configuration ensures that 22 channels are common across all EEGs within the dataset.

The EEGs were sampled at different frequencies, such as 250 Hz, 256 Hz, 400 Hz, or 512 Hz, necessitating a resampling to a common frequency prior to use. The dataset employs a hierarchical structure. Patients are distributed between a predefined training set and a predefined dev/test set. Each patient has associated recording sessions, which are further divided into smaller token files. Each token files corresponds to a time-stamped annotation file, delineating the start and end times of each seizure event, along with the seizure type.

TUSZ presents a significant challenge in the form of class imbalance. In this dataset, each seizure event is annotated according to one of eight distinct seizure classes. However, some classes are considerably underrepresented, resulting in a skewed distribution of instances across the classes. To improve training efficiency, we undersample the non-seizure samples to create equal number of samples only for the training set.

B. Experiment Results

We divided the patients from the predefined training set into training and validation sets at a ratio of 80% to 20%, respectively. Importantly, the predefined testing set did not share any patients with the predefined training set, ensuring zero data leakage at the patient level. This arrangement allows us to test our proposed methods on unseen patients, thereby ensuring the generalizability of our model.

Our deep learning framework, implemented using PyTorch, was trained on an Nvidia A100 80GB GPU. Our training parameters included a batch size of 1024, an initial learning rate of 0.0001, and the Adam optimizer. Binary cross entropy loss was used as the loss function. We trained the framework for 1000 epochs, with early stopping if no improvement in validation loss was observed over 50 consecutive epochs.

As evidenced in Table II, our approach attained a macro-average F1 score of 0.731, precision of 0.724, and recall (sensitivity) of 0.744, demonstrating the efficacy of our proposed framework in seizure detection tasks. Here, macro average is the unweighted average of the class-specific metrics.

V. Discussion

To further demonstrate the effectiveness of our proposed approach, we replicated the deep learning pipelines of several EEG analysis papers and conducted experiments on the same datasets to compare the performance. Specifically, we selected four papers published between 2018 and 2022. All experiments were conducted on the same TUSZ dataset and followed the same training process.

As shown in Table III, our proposed approach outperforms other deep learning approaches replicated from the literature.

Besides a significant class imbalance issue, another formidable challenge posed by TUSZ is the relatively low signal-to-noise ratio. As the TUSZ dataset is derived from hospital recording sessions, it undergoes considerably less preprocessing than other datasets. Consequently, these recordings often contain significant noise.

Further complicating matters, EEGs are not always obtained in standardized environments but are instead collected across multiple hospital departments. This variation in the data collection environment can introduce additional sources of noise and inconsistency in the recordings. As per the observations of other researchers, these conditions often result in models trained and evaluated on this dataset displaying lower performance metrics compared to those working with cleaner, more uniformly gathered datasets [37].

One major limitation of this work is the overfitting issue. Patient-to-patient variation is often causing the overfitting issue. Thus, adversarial learning against the patient subjects is an effective approach that can potentially mitigate overfitting, further improving seizure detection accuracy. One future direction is to include adversarial learning during training.

Another future direction is to include clinical notes as the additional data modality. From the clinical perspective, focal non-specific seizures (FNSZ) and complex partial seizures (CPSZ) have similar EEG characteristics, and the only way to

TABLE I
MODEL ARCHITECTURE AND OUTPUT SHAPE. FOR THE TRANSFORMER MODEL, WE USE 4 MSA HEADS AND 2048 AS EMBEDDING DIMENSION SIZE.

Module	Layer	# Filters	Kernel Size	Output Shape
Input EEG				(batch size, 22, 1000)
CNN	Temporal Conv2D	64	(1, K_{C1})	(batch size, 64, 22, 1000)
	Batch Norm			
	DepthWise Conv2D	256	(C, 1)	(batch size, 256, 1, 1000)
	Batch Norm			
	ELU			
	Average Pooling		(1, 5)	(batch size, 256, 1, 200)
	Separable Conv2D		(1, 16)	
	Batch Norm			
	ELU			
	Average Pooling		(1, 5)	(batch size, 256, 1, 40)
Transformer	Insert a Learnable Vector			(257, batch size, 40)
	Positional Encoding			
	Transformer Encoder Layers			
Classification	Linear Layer			(batch size, 2)

TABLE II
CLASSIFICATION REPORT OF THE PROPOSED APPROACH.

	Precision	Recall	F1-Score	Support
No Seizure	0.864	0.805	0.833	101,368
Seizure	0.584	0.683	0.630	40,650
Macro Avg	0.724	0.744	0.731	142,018
Weighted Avg	0.783	0.770	0.775	142,018

differentiate them is whether the patient remains awake when a seizure occurs. As video recordings are not available due to privacy issues, clinical notes may contain rich information about patients' status during the recording sessions. Recent studies show that transformer-based NLP models are effective in capturing clinical information from clinical notes [38], [39]. Thus, multi-modal integration of EEG features and word embeddings learned by NLP models can potentially improve seizure detection performance.

Lastly, more recently-developed transformer variants [40] and positional encoding (PE) algorithms, Scaled PE [41] or Interpolated PE [42], can be explored in future work. Explainable AI techniques [43] can also help visualize and understand deep learning classification outcomes.

VI. CONCLUSION

In this study, we introduce a novel deep learning framework designed for the automatic detection of seizures. The proposed system successfully extracts spatial and temporal seizure-specific features from raw EEG signals, while maintaining the ability to capture long-term sequential information.

We conduct extensive experiments to evaluate the performance of our proposed pipeline, utilizing a publicly accessible EEG seizure detection dataset. Our results indicate that our model's performance is comparable to, if not surpassing, that of top methodologies in the literature.

One of the main advantages of our approach is that it eliminates the need for handcrafted feature engineering and is capable of directly handling noisy EEG data, thus making it a viable option for practical clinical application. Furthermore, our methodology has demonstrated its effectiveness in detecting seizures in new and unseen patients, eliminating the requirement for patient-specific annotations.

Our work has the potential to serve as an efficient clinical decision support system for early and precise seizure detection, ultimately enhancing the quality of patient care. We remain optimistic about the impact of our research on the broader field of neurological disorder diagnosis and treatment.

REFERENCES

[1] M. M. Zack and R. Kobau, "National and state estimates of the numbers of adults and children with active epilepsy—united states, 2015," *Morbidity and Mortality Weekly Report*, vol. 66, no. 31, p. 821, 2017.

[2] I. Megiddo, A. Colson, D. Chisholm, T. Dua, A. Nandi, and R. Laxminarayan, "Health and economic benefits of public financing of epilepsy treatment in india: An agent-based simulation model," *Epilepsia*, vol. 57, no. 3, pp. 464–474, 2016.

[3] L. Hirsch, E. Donner, E. So, M. Jacobs, L. Nashef, J. Noebels, and J. Buchhalter, "Abbreviated report of the nih/ninds workshop on sudden unexpected death in epilepsy," *Neurology*, vol. 76, no. 22, pp. 1932–1938, 2011.

[4] Y. Zhu, H. Wu, and M. D. Wang, "Feature exploration and causal inference on mortality of epilepsy patients using insurance claims data," in *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*. IEEE, 2019, pp. 1–4.

[5] A. Van de Vel, K. Cuppens, B. Bonroy, M. Milosevic, K. Jansen, S. Van Huffel, B. Vanrumste, P. Cras, L. Lagae, and B. Ceulemans, "Non-eeg seizure detection systems and potential sudep prevention: state of the art: review and update," *Seizure*, vol. 41, pp. 141–153, 2016.

[6] M. Teplan *et al.*, "Fundamentals of eeg measurement," *Measurement science review*, vol. 2, no. 2, pp. 1–11, 2002.

[7] S. Noachtar and J. Rémi, "The role of eeg in epilepsy: a critical review," *Epilepsy & Behavior*, vol. 15, no. 1, pp. 22–33, 2009.

[8] H. Ronner, S. Ponten, C. Stam, and B. Uitdehaag, "Inter-observer variability of the eeg diagnosis of seizures in comatose patients," *Seizure*, vol. 18, no. 4, pp. 257–263, 2009.

[9] J. Halford, D. Shiau, J. Desrochers, B. Kolls, B. Dean, C. Waters, N. Azar, K. Haas, E. Kutluay, G. Martz *et al.*, "Inter-rater agreement on identification of electrographic seizures and periodic discharges in icu eeg recordings," *Clinical Neurophysiology*, vol. 126, no. 9, pp. 1661–1669, 2015.

[10] F. Pauri, F. Pierelli, G.-E. Chatrian, and W. W. Erdly, "Long-term eeg-video-audio monitoring: computer detection of focal eeg seizure patterns," *Electroencephalography and clinical Neurophysiology*, vol. 82, no. 1, pp. 1–9, 1992.

[11] R. S. Fisher, H. E. Scharfman, and M. DeCurtis, "How can we identify ictal and interictal abnormal activity?" *Issues in Clinical Epileptology: A View from the Bench*, pp. 3–23, 2014.

TABLE III

PERFORMANCE COMPARISON OF OUR APPROACH AGAINST THE REPLICATED STATE-OF-THE-ART DEEP LEARNING APPROACHES FROM THE LITERATURE. ALL APPROACHES ARE EVALUATED ON THE SAME TUSZ DATASET FOR BINARY SEIZURE DETECTION TASK. MSA: MULTIHEAD SELF ATTENTION.

Approach/ Paper	Key Model Components	Macro F1	Macro Precision	Macro Recall
EEGNet, 2018 [28]	EEGNet	0.700	0.706	0.696
EEG-TCNet, 2020 [29]	EEGNet + TCN	0.689	0.695	0.738
ATCNet, 2022 [30]	EEGNet + MSA + TCN	0.707	0.706	0.707
Sun et al., 2022 [24]	Shallow CNN + Transformer Encoder	0.710	0.702	0.732
Ours	**EEGNet + Transformer Encoder**	**0.731**	**0.724**	**0.744**

[12] F. Mormann, K. Lehnertz, P. David, and C. E. Elger, “Mean phase coherence as a measure for phase synchronization and its application to the eeg of epilepsy patients,” *Physica D: Nonlinear Phenomena*, vol. 144, no. 3-4, pp. 358–369, 2000.

[13] V. Srinivasan, C. Eswaran, and N. Sriraam, “Approximate entropy-based epileptic eeg detection using artificial neural networks,” *IEEE Transactions on information Technology in Biomedicine*, vol. 11, no. 3, pp. 288–295, 2007.

[14] Y. Zhang, S. Yang, Y. Liu, Y. Zhang, B. Han, and F. Zhou, “Integration of 24 feature types to accurately detect and predict seizures using scalp eeg signals,” *Sensors*, vol. 18, no. 5, p. 1372, 2018.

[15] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. Abd El-Samie, “Eeg seizure detection and prediction algorithms: a survey,” *EURASIP Journal on Advances in Signal Processing*, vol. 2014, pp. 1–21, 2014.

[16] A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos, and S. Van Huffel, “Neonatal seizure detection using deep convolutional neural networks,” *International journal of neural systems*, vol. 29, no. 04, p. 1850011, 2019.

[17] F. Achilles, F. Tombari, V. Belagiannis, A. M. Loesch, S. Noachtar, and N. Navab, “Convolutional neural networks for real-time epileptic seizure detection,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, vol. 6, no. 3, pp. 264–269, 2018.

[18] X. Zhang, L. Yao, M. Dong, Z. Liu, Y. Zhang, and Y. Li, “Adversarial representation learning for robust patient-independent epileptic seizure detection,” *IEEE journal of biomedical and health informatics*, vol. 24, no. 10, pp. 2852–2859, 2020.

[19] A. M. Abdelhameed, H. G. Daoud, and M. Bayoumi, “Deep convolutional bidirectional lstm recurrent neural network for epileptic seizure detection,” in *2018 16th IEEE International New Circuits and Systems Conference (NEWCAS)*. IEEE, 2018, pp. 139–143.

[20] M. Saqib, Y. Zhu, M. Wang, and B. Beaulieu-Jones, “Regularization of deep neural networks for eeg seizure detection to mitigate overfitting,” in *2020 IEEE 44th Annual Computers, Software, and Applications Conference (COMPSAC)*. IEEE, 2020, pp. 664–673.

[21] Y. Li, Z. Yu, Y. Chen, C. Yang, Y. Li, X. Allen Li, and B. Li, “Automatic seizure detection using fully convolutional nested lstm,” *International journal of neural systems*, vol. 30, no. 04, p. 2050019, 2020.

[22] C. Lea, R. Vidal, A. Reiter, and G. D. Hager, “Temporal convolutional networks: A unified approach to action segmentation,” in *Computer Vision–ECCV 2016 Workshops: Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 47–54.

[23] E. Aksan and O. Hilliges, “Stcn: Stochastic temporal convolutional networks,” in *7th International Conference on Learning Representations (ICLR 2019)*, 2019.

[24] Y. Sun, W. Jin, X. Si, X. Zhang, J. Cao, L. Wang, S. Yin, and D. Ming, “Continuous seizure detection based on transformer and long-term ieeg,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 11, pp. 5418–5427, 2022.

[25] J. Pedoeem, G. Bar Yosef, S. Abittan, and S. Keene, “Tabs: Transformer based seizure detection,” in *Biomedical Sensing and Analysis: Signal Processing in Medicine and Biology*. Springer, 2022, pp. 133–160.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.

[27] C. Li, X. Huang, R. Song, R. Qian, X. Liu, and X. Chen, “Eeg-based seizure prediction via transformer guided cnn,” *Measurement*, vol. 203, p. 111948, 2022.

[28] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, “Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces,” *Journal of neural engineering*, vol. 15, no. 5, p. 056013, 2018.

[29] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, “Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces,” in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020, pp. 2958–2965.

[30] H. Altaheri, G. Muhammad, and M. Alsulaiman, “Physics-informed attention temporal convolutional network for eeg-based motor imagery classification,” *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 2249–2258, 2022.

[31] Y. Zhu, M. Saqib, E. Ham, S. Belhareth, R. Hoffman, and M. D. Wang, “Mitigating patient-to-patient variation in eeg seizure detection using meta transfer learning,” in *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. IEEE, 2020, pp. 548–555.

[32] P. Thuwajit, P. Rangpong, P. Sawangjai, P. Autthasan, R. Chaisaen, N. Banluesombatkul, P. Boonchit, N. Tatsaringkansakul, T. Sudhawiyangkul, and T. Wilaiprasitporn, “Eegwavenet: Multiscale cnn-based spatiotemporal feature extraction for eeg seizure detection,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 8, pp. 5547–5557, 2021.

[33] R. Peng, C. Zhao, J. Jiang, G. Kuang, Y. Cui, Y. Xu, H. Du, J. Shao, and D. Wu, “Tie-eegnet: Temporal information enhanced eegnet for seizure subtype classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2567–2576, 2022.

[34] K. Cho, B. Van Merriënboer, D. Bahdanau, and Y. Bengio, “On the properties of neural machine translation: Encoder-decoder approaches,” *arXiv preprint arXiv:1409.1259*, 2014.

[35] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 2019, pp. 4171–4186.

[36] I. Obeid and J. Picone, “The temple university hospital eeg data corpus,” *Frontiers in neuroscience*, vol. 10, p. 196, 2016.

[37] M. Golmohammadi, S. Ziyabari, V. Shah, S. L. de Diego, I. Obeid, and J. Picone, “Deep architectures for automated seizure detection in scalp eegs,” *arXiv preprint arXiv:1712.09776*, 2017.

[38] Y. Zhu, A. Mahale, K. Peters, L. Mathew, F. Giuste, B. Anderson, and M. D. Wang, “Using natural language processing on free-text clinical notes to identify patients with long-term covid effects,” in *Proceedings of the 13th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics*, 2022, pp. 1–9.

[39] K. Mermin-Bunnell, Y. Zhu, A. Hornback, G. Damhorst, T. Walker, C. Robichaux, L. Mathew, N. Jaquemet, K. Peters, T. M. Johnson *et al.*, “Use of natural language processing of patient-initiated electronic health record messages to identify patients with covid-19 infection,” *JAMA Network Open*, vol. 6, no. 7, pp. e2 322 299–e2 322 299, 2023.

[40] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong *et al.*, “Swin transformer v2: Scaling up capacity and resolution,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 12 009–12 019.

[41] N. Li, S. Liu, Y. Liu, S. Zhao, and M. Liu, “Neural speech synthesis with transformer network,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 6706–6713.

[42] G. Bertasius, H. Wang, and L. Torresani, “Is space-time attention all you need for video understanding?” in *ICML*, vol. 2, no. 3, 2021, p. 4.

[43] F. Giuste, W. Shi, Y. Zhu, T. Naren, M. Isgut, Y. Sha, L. Tong, M. Gupte, and M. D. Wang, “Explainable artificial intelligence methods in combating pandemics: A systematic review,” *IEEE Reviews in Biomedical Engineering*, 2022.